ChinaXiv合作期刊

图书情报工作
第 62 卷 第 13 期　2018 年 7 月

基于开放同行评议的学者影响力评价研究*
——以 F1000 为例

袁国华[1,2]　寇晶晶[1,2]　张建勇[1]　韩正琪[1,2]

[1] 中国科学院文献情报中心　北京 100190　[2] 中国科学院大学　北京 100049

摘要：[**目的/意义**]提出一种基于开放同行评议的学者影响力评价方法，即通过归纳收敛开放同行评议内容，形成能够表征学者不同层面影响力的多个标签簇，从而从质量层面多角度探索学者影响力，并验证所提方法的有效性。[**方法/过程**]调研分析国内外现有学者影响力评价方法，在此基础上，设计构建基于开放同行评议的学者影响力评价方法，并以开放同行评议平台 F1000 为例，进行实证研究和对比分析。[**结果/结论**]基于开放同行评议的学者评价方法是一种新的评价角度，实现了对学者的多角度评价，增加了学者影响力的饱满度和辨识度。分析表明，该方法与文献计量评价方法不具有显著相关性，可以作为对传统评价方法的一种补充。传统的定量分析与开放同行评议方法的结合将是未来学术评价的发展趋势。

关键词：学者影响力　评价　开放同行评议　F1000

分类号：G250

DOI：10.13266/j.issn.0252-3116.2018.13.006

chinaXiv:202308.00646v1

1 引言

学者影响力评价对科研管理和评估起着非常重要的作用，一直以来都是学术评价领域关注的重点。有效的学者影响力评价有利于促进科学研究的公平竞争，提升学者从事科学研究的积极性。学者影响力的广度和深度主要取决于其研究的成果以及这些成果被他人重视、认同和应用的情况，因此，对学者的评价，归根结底是对学者学术成果的评价。现有的学术评价主要依赖文献计量学的方法，其中基于引文分析的方法使用最为广泛，这种评价方法在得到广泛认可的同时，也不断遭到诟病。一直以来，科研人员都希望在现有计量评价方法的基础上，辅以定性评价的方法。尤其是在科学出版、传播与交流方式日益开放化的今天，社交媒体的蓬勃发展为开放式的评价提供了更加合适的平台和媒介，开放同行评议（open peer review，OPR）应运而生，海量的定性评价成为可能。

2017 年 5 月，欧盟 OpenAIRE 项目发布了名为“OpenAIRE 调查编辑、作者、评议者对开放同行评议的态度与体验”（OpenAIRE survey on open peer review：attitudes and experience amongst editors，authors and reviewers）的调查报告[1]，调查结果显示，有四分之三的受访者表示他们以不同的方式或多或少地参与过开放同行评议。开放同行评议正以多种形式渗透于科研评价中。鉴于此，本文构建了基于开放同行评议的学者影响力评价方法和流程，通过归纳收敛开放同行评议内容，形成了 3 种不同的标签簇，代表学者在不同方面影响力，从而从质量层面多角度展开对学者影响力的评价，增强学者之间的辨识度，并以 F1000 作为案例展开实证分析研究，验证所提方法的有效性。

2 国内外研究进展

2.1 现有学者影响力评价方法

以 18 世纪中叶爱丁堡皇家学会和伦敦皇家学会邀请其成员挑选用于出版的文章为起点，论文同行评议制度至今已有近 300 年的历史[2]。期间，学术评价经历了从同行评议为主，到同行评议与文献计量并存，再到文献计量为主的发展历程。在学者影响力评价领

* 本文系国家科技图书文献中心项目“名称规范库建设”（项目编号：科 1536）研究成果之一。

作者简介：袁国华（ORCID：0000-0002-4881-0155），工程师，博士研究生；寇晶晶（ORCID：0000-0002-1713-4828），硕士研究生，通讯作者，E-mail：koujingjing@mail.las.ac.cn；张建勇（ORCID：0000-0001-7533-1726），研究馆员，硕士生导师；韩正琪（ORCID：0000-0002-7998-9717），硕士研究生。

收稿日期：2018-03-08　修回日期：2018-04-08　本文起止页码：37-44　本文责任编辑：刘远颖

ChinaXiv合作期刊

域,1926年,洛特卡在《科学生产率的频率分布》中提出了描述著者与论文数量关系的经验规律,后被称为"洛特卡定律"[3],为将学者产出与学者影响力评价联系在一起奠定了基础。1969年,"文献计量学"一词被正式提出,国内外基于文献计量学对学者影响力进行评价的方法层出不穷,从加菲尔德提出的被引频次评价到J. Hirsch提出的h指数[4],再到多种h指数的变体、PageRank变体等,科研人员在学者影响力评价领域进行了坚持不懈的探索,如L. Egghe提出的g指数[5],C. T. Zhang[6]提出的e指数,国内学者提出的Author-Rank[7]、LeaderRank[8]等。具体包括均值型指标、h型指标和关联型指标[9]3类。

均值型指标以发文、引文的均值作为核心思想,主要包括篇均被引频次、影响因子,试图以均值的方法消除文章数量或年份带来的影响,但并没有达到理想的效果,如篇均被引频次建立在被引的基础上,就忽略了引用动机对评价结果的影响。

h型指标是以h指数为基础演化出的一系列指标,学术界对h指数的研究可谓经久不衰,西班牙还有专门针对h指数相关研究的网站"H-index and Variants"[10],提供了包括h指数的原理、应用和h指数变体等多方面的内容。据不完全统计,h型指标目前有30多个成员,分别针对h指数的不同缺陷进行了不同程度的改进。

关联型指标是指依据学者之间存在的关联关系对学者进行评价的指标,具体指作者之间的相互引用和合作关系。这类指标中较为热点的研究是对PageRank的改进,利用网页链接与作者引用、合作关系之间的相似性,产生出了多种关联型指标,如孙海生等提出的AuthorRank、王小梅等提出的LeaderRank、马瑞敏等提出的使用特征因子评价学者的方法[11]等。

以文献计量学为背景的学者影响力评价方法是目前使用最广泛、最成功的评价方法,但关于该类方法弊端的讨论也从未停止过。一方面体现在其比较注重定量的评价,而忽略了从定性层面审视文章的影响力;另一方面现有方法的评价角度略显单一,每个作者在一种评价方法下只能得到唯一的得分,而没有体现出作者的个人特征,这种一刀切的做法,需要得到进一步的改进和补充。

2.2 开放同行评议的产生与发展

随着信息技术和社交媒体平台的不断发展,学术成果出版、传播、交流的方式也与时俱进,学术成果的获取和评价开始转向网络和社交媒体,这为开放的评价提供了合适的平台和媒介,开放同行评议用于学术评价逐渐成为可能。

开放同行评议,是同行评议在开放科学时代的一种体现形式,"开放同行评议"一词最早出现在20世纪末,起初,这一方式作为对传统同行评议的一种补充被提出,之后,渗透到出版中和出版后的过程中。开放同行评议的这种发展历程,也使得学术界对它的定义众说纷纭,如N. McCormack[12]认为开放同行评议是作者和审稿人身份互知;A. Mulligan[13]则认为既要作者和审稿人身份互知,同时他们的身份也向大众披露。D. Shotton[14]和P. Perakakis[15]认为开放同行评议是一个大众参与决定稿件去留的过程,即稿件收稿后立即接受公开的评论,这些评论与正式同行评议结果一并决定稿件去留。国内学者刘春丽[16]将之定义为在网络社区中,由同行评议专家根据自己专长的研究领域,快速鉴别学术研究出版物中重要文献的过程。姜春林[17]认为开放同行评议是同行专家利用社交工具对开放存取平台上的论文在线评审,同时论文作者与同行专家进行互动交流的过程。

E. Ford[18]对其进行了梳理,给出了开放评议的8个特征,其中5个特征是描述评议过程的,包括评审意见是否署名、评审过程中作者与评审人的身份是否相互公开并可相互讨论、编辑是否控制评审过程、评议结果是否向大众公开、是否允许读者评议。另外3个是描述评议时间的,包括出版前评议、出版中评议、出版后评议。

本文的研究是基于出版后的开放同行评议展开的,在本文当中,出版后开放评议专指经过传统同行评议认可后出版的文章,再次接受其他同行或读者评议的过程。

开放同行评议的产生,激发了学术评价领域研究的热情。目前,国内外已经有一部分学者尝试了使用开放同行评议对文章、学者等展开评价,并与以引文为主的文献计量评价方法进行了相关性比较研究。如宋丽萍等基于F1000和WoS比较了同行评议与文献计量方法在科学评价中的有效性及相关性[19],得出了文献计量指标与同行评议结果具有正向相关性的结论。沈小玲等[20]通过探索与同行评议相关联的网络计量指标,以此替代同行评议建立了综合评价模型。

而笔者通过对利用开放同行评议进行学术评价的现有文章进行调研发现,这些研究大多使用开放同行评议的定量指标,如F1000因子、Altmetrics指标等,但同时,开放同行评议还会产生定性的评议内容,这些评

chinaXiv:202308.00046v1

袁国华，寇晶晶，张建勇，等.基于开放同行评议的学者影响力评价研究——以 F1000 为例[J].图书情报工作,2018,62(13):37－44.

议内容对学术评价来说也非常重要。因此,不同于以往的研究,本文将视角放在了开放同行评议产生的定性内容上,将以 F1000 标签作为学者影响力评价的依据,构建基于开放同行评议的学者影响力评价方法。

3 基于开放同行评议的学者影响力评价方法构建

基于开放同行评议的学者影响力评价方法的核心思想是:通过归纳收敛开放同行评议内容,形成能够表征学者不同层面影响力的多个标签簇,从而从质量层面、多角度探索学者影响力。由于文章的评价方法以 F1000 为基础,因此,有必要首先对 F1000 的开放同行评议情况进行详细的说明。

3.1 F1000 开放同行评议情况简介

F1000,全称 Faculty of 1000[21],是出版后开放同行评议的一个典型代表。F1000 对自身的定义是为生命科学领域的文章提供二次评估的数据库。F1000 的开放评议以 3 种形式存在,分别为打分制、文本评议和标签评议,参加评议的成员为美国和欧洲等国际知名机构的著名专家,他们根据论文对当前世界生物和医学研究的贡献程度和科学价值给出评价。打分制是指评议专家将文章按 Good、Very good 和 Exceptional 打分,分别代表一星、二星、三星,所有评议专家的打分综合起来就得到了文章的 F1000 因子。并根据 F1000 因子将最近一个月内的极少数优秀论文推荐给读者。文本评议是指专家以文本的形式写出对文章的评价意见。而标签评议是指专家给出文章评议标签,这里的标签不是任意词语都可以,F1000 提供了标签库[22],如表 1 所示，评议专家需要从标签库中选择任意一个或多个

chinaXiv:201806.00616v1

表 1 F1000 中的标签及其释义

标签	标签释义
Confirmation	证实之前公布的数据或假设
Controversial	挑战已有的权威
Interesting Hypothesis	提出新的模型
New Finding	新发现,提供新的数据、模型或假设
Technical Advance	介绍一种新的实践/理论技术,或现有技术的新应用
Good for Teaching	领域的重要文章或模板
Refutation	否定之前公布的数据或假设
Novel Drug Target	为药物发现新靶点
Changes Clinical Practice	完整的临床实践
Systematic Review/Meta-analysis	系统回顾,荟萃分析
Review/Commentary	综述/评论性的文章

标签来表示对文章的评价。以 Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis 一文[23]为例,图 1 是专家 J. Wall 对该文的评议,可以看到他对这篇文章的评价很高,给出了三星的打分,标签是 new finding,同时给出了三段文本评议。

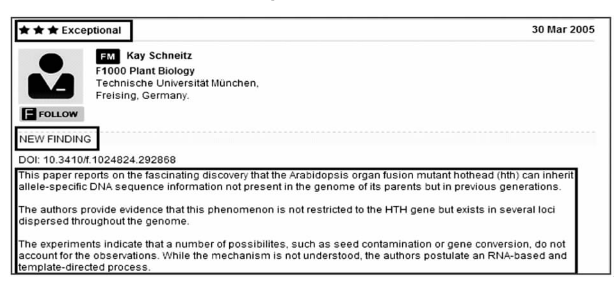


图 1 F1000 评议示意

在本文中,笔者选用 F1000 作为开放同行评议的代表,构建基于开放同行评议的学者影响力评价方法,其原因如下:

(1)F1000 不以期刊作为文章和作者评价的依据,既关注人们普遍认为的高质量期刊上的文章,也关注普通期刊上的文章,对所有的期刊一视同仁。到目前为止,F1000 已经推荐了 3 500 多种期刊的文章。相较于以往的评价方法,F1000 提供了一种更公平的评价方法,而不是只关注高质量期刊上的文章或作者,忽略那些文章质量高但缺乏知名度的作者。

(2)F1000 中评议结果的专业度高。F1000 汇集了 8 000 多位全球生命科学领域最顶尖的科学家,既保证了开放评议的广泛性,也保证了评议同行的专业性,评议结果可信赖度较高,从而让基于这一平台的开放同行评议评价方法具有较多可利用的数据。

(3)F1000 中的开放评议数据较为规范化。F1000 的评议结果主要以评议星级和评议内容两种方式呈现。评议星级最终构成了 F1000 因子,F1000 因子是文章评价最直接的数据依据,F1000 因子越高,表示文章受专家推荐等级越高。而评议内容主要以文本和标签两种形式体现,二者是相关关系,标签是对文本评议内容的结构化。

(4)F1000 中的评议内容较丰富,具有较多的可利用内容。虽然网络和社交媒体正处于飞速发展的时代,但学术平台的开放评议还处于初级阶段,很多平台虽然具有开放评议的功能,但是评议数量和内容非常少,而 F1000 不同,它在开放的同时,带有一定的强制性,这也使得其评议数量和内容较丰富,适合做基于开放评议数据的学术评价。

ChinaXiv合作期刊

3.2 评价思路构建

在进行学者评价时,不论是传统的评价方法,还是新型的评价方法,都是以作者的研究成果,或者更确切地说是作者论文为依据展开的,本文的评价方法也不例外。图 2 为依据上述核心思想,以 F1000 及其学者论文为依据构建的基于开放同行评议的学者影响力评价思路。具体而言,主要涉及以下几个步骤:

(1)信息抓取:以 F1000 为基础,抓取 F1000 上学术论文的作者、题名、标签、F1000 因子等信息。

(2)标签统计:对文章的标签进行统计,得到每篇文章拥有的每个标签数量。

(3)标签簇划归:由于 F1000 上的标签数量较多,为了更方便分析和计算,也更准确地突出作者的特征,参考期刊文章划分,将标签划归为不同类簇,表征不同的特征,在本文中称为标签簇。具体的整合结果见表 2。

(4)标签迁移:分离出每篇文章的第一作者,将标签簇从所属文章迁移到对应的学者上,形成作者的标签簇,在本文中称为 Label Score,从而表征作者不同方面的特征。

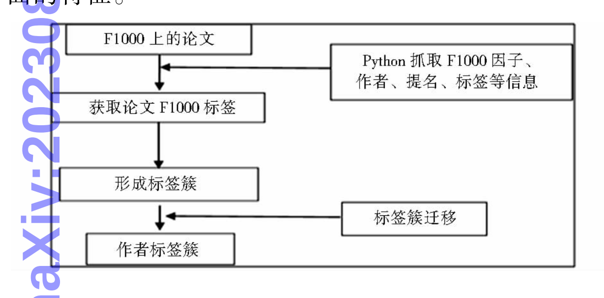


图 2 基于开放同行评议的学者影响力评价思路

(5)实证研究:以 F1000 为基础进行实证研究,并与已有评价方法做相关性分析和比较分析。

表 2 标签整合结果

标签	标签簇
New Finding	New Finding
Interesting Hypothesis	
Controversial	Clinical Practice
Confirmation	
Novel Drug Target	
Changes Clinical Practice	
Refutation	
Review/Commentary	Review
Systematic Review/Meta-analysis	
Technical Advance	
Good for Teaching	

4 实证研究

4.1 数据采集

本次实验数据来源于 F1000 和 WoS。将 F1000 作为开放同行评议的代表,选取心血管学科,使用 Python 爬虫工具得到该学科 4 079 篇文章的作者、题名、F1000 因子、标签等信息(数据爬取时间:2017 年 12 月 7 日),保存为. csv 文档,部分结果见图 3。WoS 作为世界权威的文献索引数据库,收录范围广,能够直观地从引文角度评估论文的学术价值,可以作为现有主流评价方法——文献计量学的代表。

序号	title	authors	score	time	lable
0	We must f	Ringach DL,Jent	48	2009	Confirmation,Controversial,General Interest,Confirmation,Controversial,General Interest,,Interesting Hypothesis,New Finding,Controversial,General Inte
1	Transgenic	Livet J,Weissma	38	2007	New Finding,Technical Advance,Technical Advance,Technical Advance,Technical Advance,New Finding,Technical Advance,Technical Advance,New Find
2	Structural	Chung K,Wallac	35	2013	Technical Advance,New Finding,Technical Advance,Technical Advance,Good for Teaching,Technical Advance,Technical Advance,New Finding,Technica
3	Mortality a	Maitland K,Kigu	26	2011	New Finding,Controversial,New Finding,Controversial,New Finding,New Finding,New Finding,Refutation,Controversial,New Finding,Controversial,New
4	Crystal stru	Rasmussen SG,I	26	2011	New Finding,Technical Advance,New Finding,Technical Advance,New Finding,Technical Advance,New Finding,New Finding,Novel Drug Target,Technic
5	NIH peer r	Fang FC,Bowen	25	2016	Controversial,General Interest,,Good for Teaching,Controversial,General Interest,,Good for Teaching,Refutation,Controversial,General Interest,Controve
6	The crystal	Payandeh J,Sch	25	2011	New Finding,Technical Advance,New Finding,Technical Advance,New Finding,New Finding,Novel Drug Target,Technical Advance,New Finding,New Fin
7	K+ channe	Choi M,Scholl L	24	2011	New Finding,New Finding,New Finding,Novel Drug Target,New Finding,Novel Drug Target,New Finding,New Finding,New Finding
8	Rescuing L	Alberts B,Kirsch	21	2014	Controversial,General Interest,,Good for Teaching,Interesting Hypothesis,Review / Commentary,General Interest,,Interesting Hypothesis,Review / Comr
9	Circulating	Zhang Q,Raoof	21	2010	Interesting Hypothesis,New Finding,Novel Drug Target,Interesting Hypothesis,New Finding,New Finding,Interesting Hypothesis,New Finding,New Findi
10	Macropha	Machnik A,Neul	21	2009	Interesting Hypothesis,New Finding,New Finding,New Finding,Novel Drug Target,Interesting Hypothesis,New Finding,New Finding,Novel Drug Target,I
11	Induction	Wu C,Yosef N,T	20	2013	New Finding,Interesting Hypothesis,New Finding,Novel Drug Target,Interesting Hypothesis,New Finding,Good for Teaching,New Finding,Novel Drug T
12	Statistical t	Greenland S,Ser	19	2016	Confirmation,General Interest,,Good for Teaching,Confirmation,Controversial,General Interest,,Good for Teaching,Controversial,General Interest,,Good
13	Beyond ba	Weissgerber TL,	19	2015	General Interest,,Good for Teaching,Systematic Review / Meta-analysis,General Interest,,Good for Teaching,Confirmation,General Interest,,Good for Te
14	Kindlin-3 i	Moser M,Niesw	19	2008	New Finding,Novel Drug Target,Interesting Hypothesis,New Finding,New Finding,New Finding,Novel Drug Target,New Finding,New Finding,Interesting
15	Gut flora n	Wang Z,Klipfell	18	2011	New Finding,Novel Drug Target,Technical Advance,New Finding,New Finding,Novel Drug Target,Good for Teaching,New Finding,New Finding,New Fin
16	Transvascu	Kumar P,Wu H,I	18	2007	New Finding,Technical Advance,New Finding,Technical Advance,New Finding,Technical Advance,New Finding,Technical Advance,New Finding,Technic

图 3 爬取的部分数据示例

4.2 数据处理

首先对爬取的数据做如下处理:

(1)利用 Excel 和 Python 将每篇文章得到的每个标签数量进行汇总。

(2)利用整合过的标签簇,对每篇文章的标签进行二次求和汇总,得到每篇文章的标签簇,部分数据见图 4。

(3)采用作者贡献标准分配法,即将文章的所有贡献都分配给第一作者的方法进行实验。在上一步的基础上将文章的第一作者提出。

chinaXiv:202308.00646v1

ChinaXiv合作期刊

（4）按照第一作者姓名进行分类汇总，得到发文量前 20 的作者。

（5）求得每位作者的特征值。每位作者在每个标签簇上的得分称为作者特征值，在本文中用 Label Score 表示。特征值的计算方法可以用公式（1）表示。N 表示作者在 F1000 上拥有的文章数量，i 表示文章，i＝1，2，…，N，j 表示标签簇，j＝1，2，3，a_{ij} 表示作者的文章 i 在标签簇 j 上的得分。这样，每个作者就可以得到 4 个表示其特征的 Label Score。

$$\text{Label Score} = \frac{\sum_{i=1}^{N} a_{ij}}{N} \qquad \text{式(1)}$$

（6）找到每位作者对应的文章，手动在 WoS 中检索这些文章，并记录其被引频次，同时计算每位作者的篇均被引频次，作为文献计量方法评价的依据。

文章	Confirmat	Controv	General	Interest	New Fin	Technical	Good	Refutatio	Novel Dr	Review	Systematic Rev	Clinical Tri	Negative / Null Results
文章0	5	16	17	2	2								
文章1					3	16	1						
文章2					4	12	2						
文章10				3	8				3				
文章100	1				3				3				
文章1000	1				1								
文章1001						1	1						
文章1002	1				1					1			
文章1003				1									
文章1004				1	1	1							
文章1005	1			1		1	1						
文章1006				1	1								
文章1007				1			1						
文章1008	1				1	1	1		1	1			
文章1009				1	1								
文章101					2	2							
文章1010									1	1			

图 4　文章及标签对应情况

4.3　实证结果与分析

表 3 为使用本文提出的评价方法对发文量居前 20 位的作者的评价结果，将表 3 进行可视化，结果见图 5。图 5 中横坐标表示 Label Score，纵坐标为作者姓名，图中不同颜色条代表不同的 Label Score，颜色条的长短表示 Label Score 的高低，当作者在某一标签上得到的 LabelScore 较高时，对应的颜色条长度就会较长，相反则较短。结合表 3 和图 5 可以得出，本文评价方法中每位学者不再得到单一评价得分，而是以多个得分反映学者多方面的特征，得到更为全面的评价结果。

表 3　作者标签簇、篇均 F1000 因子及篇均被引频次

作者＼得分	Label Score			篇均被引频次	篇均 F1000 因子
	Clinical Practice	New Finding	Review		
J. Lin	0.33	0.67	0.33	36.67	5.33
H. Zhang	0	0.67	0	24	1.50
M. Amann	0	0.67	0	95	1.00
J. Liu	0.20	0.67	0.20	12	4.33
J. J. McMurray	0.33	0.67	0	1274.33	1.67
S. P. Mortensen	0	1.00	0	57.33	1.67
C. O. Tan	0.33	1.00	0.33	22.67	1.67
F. Liu	0	1.67	0	84	2.00
G. Grassi	0	1.00	0	9.33	1.00
L. Li	0.50	0.75	0	48.33	1.50
B. Pitt	0.25	0.75	0	648.75	2.00
Y. Lu	0.33	1.00	0.33	38.25	1.75
Y. Liu	0.33	1.25	0	24	1.33
X. Wang	0.17	1.50	0.17	45.75	1.00
X. Zhang	0	0.80	0	174.8	2.17
Z. Chen	0	1.00	0	15	1.50
J. Li	0.25	1.00	0	47.75	1.33
Y. Zhang	0	1.00	0	49.33	1.25
D. Chen	0	1.00	0	26.67	1.33
M. A. Laflamme	0.33	1.00	0.67	537	1.67

chinaXiv:202308.00646v1

ChinaXiv合作期刊

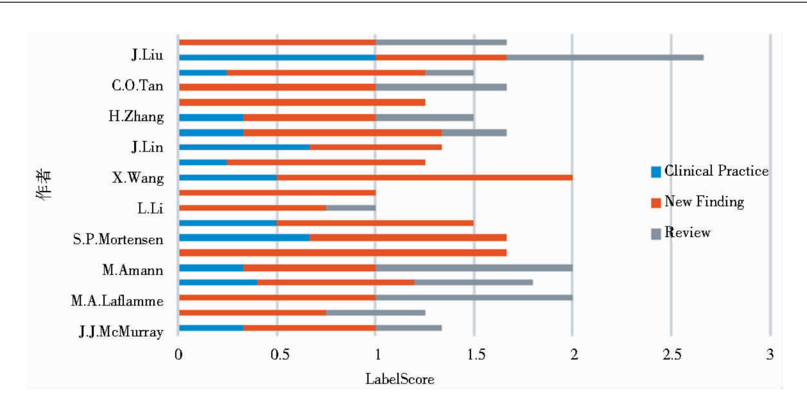


图 5 作者在不同标签簇上的表现(Label Score)

(1)横向方面,从作者在不同标签上的得分和颜色条长度,可以对作者的特征进行大致的判定。为能更加直观地说明情况,在图 6 中,给出了 J. J. McMurray、J. Liu、M. Amann 3 位作者的标签簇得分雷达图,从雷达图中每条线的分布情况可以更清晰地观测到 3 个作者在不同标签簇上的表现情况。J. J. McMurray 在"New finding"表现突出,J. Liu 在"Clinical Practice"和"Review"上的表现相对高,而 M. Amann 的特征则主要体现在"Review"上。

(2)纵向方面,可以得知在同一标签上哪些作者的表现比较突出,如在实践方面 X. Wang 相对其他作者要突出一些,在创新方面 F. Liu 要突出一些,J. Liu 存在的争议性更大一些,而 J. Liu、M. A. Laflamm 和 M. Amann 在论述方面比较突出,对应的颜色条要长一些。这种纵向的对比有助于快速地获得某一特征上表现突出的作者和对应的杰出文章。

4.4 对比分析

为了更好地验证本文评价方法的有效性和必要性,

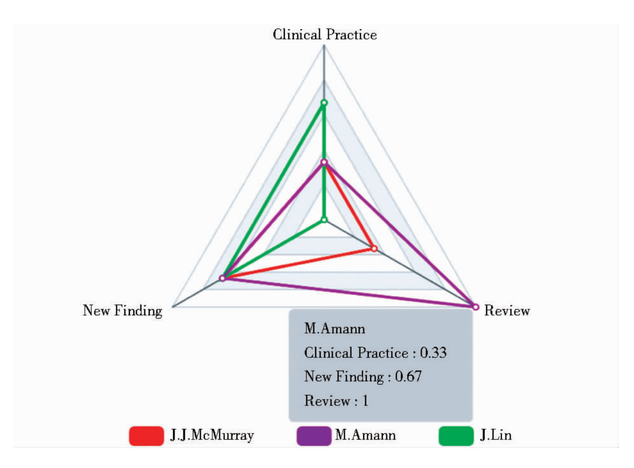


图 6 部分作者标签簇得分雷达图

在现有基础上将开放同行评议的结果与篇均被引频次代表的文献计量评价方法做相关性比较分析。首先利用 Excel 得到表 3 中数据的趋势图(见图 7),之后采用 SPSS 软件对开放同行评议的标签簇、篇均 F1000 因子和篇均被引频次做皮尔逊双侧相关性检验,结果见表 4:

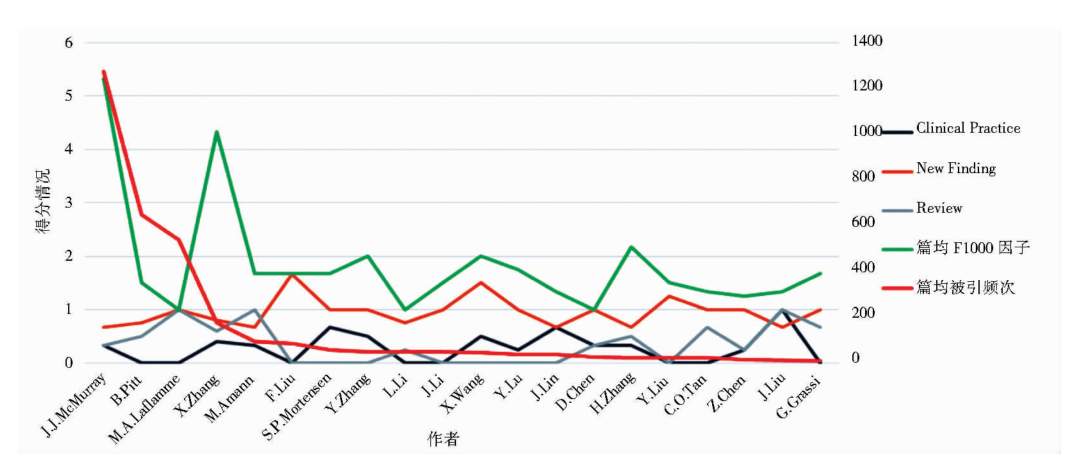


图 7 各标签、篇均 F1000 因子及被引频次趋势

chinaXiv:202303.00640v1

袁国华，寇晶晶，张建勇，等. 基于开放同行评议的学者影响力评价研究——以 F1000 为例[J]. 图书情报工作,2018,62(13):37－44.
ChinaXiv合作期刊

表 4　各标签簇、F1000 因子及篇均被引频次相关性检验

		Clinical Practice	New Finding	Review	篇均被引频次	篇均 F1000 因子
Clinical Practice	Pearson 相关性	1	－.301	.023	－.135	.153
	显著性(双侧)		.197	.923	.569	.520
	N	20	20	20	20	20
New Finding	Pearson 相关性	－.301	1	－.475*	－.275	－.209
	显著性(双侧)	.197		.034	.241	.376
	N	20	20	20	20	20
Review	Pearson 相关性	.023	－.475*	1	.178	－.005
	显著性(双侧)	.923	.034		.452	.982
	N	20	20	20	20	20
篇均被引频次	Pearson 相关性	－.135	－.275	.178	1	.625**
	显著性(双侧)	.569	.241	.452		.003
	N	20	20	20	20	20
篇均 F1000 因子	Pearson 相关性	.153	－.209	－.005	.625**	1
	显著性(双侧)	.520	.376	.982	.003	
	N	20	20	20	20	20

注:* 表示在 0.05 水平(双侧)上显著相关;** 表示在 .01 水平(双侧)上显著相关

结合图 7 和表 4 可以看出:

(1)篇均 F1000 因子与篇均被引频次之间的相关系数为 0.625,在 0.01 水平(双侧)上显著相关,图 7 中二者的基本趋势也类似,表明开放同行评议的定量评价指标 F1000 与文献计量的评价指标篇均被引频次具有较高的一致性,这为验证开放同行评议在学术评价中的有效性提供了依据。

(2)Clinical Practice、New Finding 和 Review 3 种标签簇与篇均被引频次的相关系数分别为 －0.135、－0.275、0.178,同时结合对应的 Sig 值以及各标签簇与篇均被引频次趋势折线图(见图 7),可以看出这 3 种标签簇与篇均被引频次的相关关系非常弱,甚至不存在相关关系。这在一定程度上可以说明,基于开放同行评议的定性评价涉及了传统评价方法尚未涉及的层面,重点关注的作者也与传统评价方法下的知名作者有所不同。

(3)3 种标签簇与 F1000 因子的相关系数分别为 0.153、－0.209、和 0.005,同时结合对应的 Sig 值看出,3 种标签簇与 F1000 因子的相关性也不显著,这表明我们在关注各开放同行评议平台、各社交媒体的定量指标的同时,也应考虑其定性的评价内容。

综上,本文认为,基于开放同行评议的学者评价方法是一种新的评价角度,可以作为对传统评价方法的一种补充。

5　结语

本文调研了国内外现有学者影响力评价方法,并进行了系统分类,之后在此基础上,提出一种基于开放同行评议的学者影响力评价方法,并以开放同行评议平台 F1000 为例,进行了实证研究和对比分析。在实证研究过程中,从对样本的观察可以看出,一些被引频次高的作者在开放同行评议角度下表现并不突出,而一些被引频次并不高的作者,反而在开放同行评议角度下表现突出,这在一定程度上说明了二者之间的互补关系。而通过进一步的实验分析和相关分析,本文得出如下结论:

(1)本文提出的基于开放同行评议学者影响力评价方法可以为学者赋予多重特征,打破了学者影响力评价角度单一的困境,实现了多角度的学者影响力评价,让学者的个人形象更具辨识度和饱满度,学者影响力也更加全面、生动。

(2)从统计学角度来看,基于 F1000 的开放同行评议方法与引文分析为代表的文献计量学评价方法不具有显著相关性。某种程度上表明,基于开放同行评议内容的定性评价涉及到了传统评价方法尚未涉及的层面,重点关注的作者也与传统的评价方法下的知名作者有所不同,本文研究结果为使用开放同行评议进行学者影响力评价提供了一定的参考依据。

(3)基于开放同行评议的学者评价方法是一种新的评价角度,可以作为对传统评价方法的一种补充。学者评价方法并非非此即彼,需要进行多元的结合才能更全面地评价学者。传统的评价方法与开放同行评议评价方法是相辅相成的,并非是要用开放同行评议的方法取代传统评价方法。无论是开放同行评议还是文献计量,单独作为学者评价标准都有失偏颇,传统的定量分析与开放同行评议方法的结合将是未来学术评价的发展趋势。

本文的研究也存在不足之处,研究中并未全面考虑关于作者署名与贡献度归属的问题,只对第一作者的开放同行评议情况进行了研究,这在一定程度上会影响实验结果。要真正做好基于开放同行评议的学者影响力评价,将来还需深入研究署名顺序和贡献度因素对学者影响力的影响。

chinaXiv:202308.00646v1

参考文献:

[1] Open AIRE. Open AIRE survey on open peer review: attitudes and experience amongst editors. authors and reviewers [EB/OL]. [2018 - 01 - 22]. https://zenodo. org/record/570864 #. wzvrz49oloy.
[2] KRONICK D A. Peer review in 18th-century scientific journalism. [J]. Jama the journal of the American Medical Association, 1990, 263(10):1321 -1322.
[3] 张海燕.基于文献计量学的洛特卡定律研究综述[J]. 中华医学图书情报杂志. 2013,22(8):18 -21.
[4] HIRSCH J. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy Of Sciences, 2005,102(46):16569 -16572.
[5] EGGHE L. An improvement of the h-index: the g-index[J]. ISSI newsletter,2006,2(1):8 -9.
[6] ZHANG C T. The e-index, complementing the h-index for excess citation[J]. PLoS ONE,2009,4(5):1 -4.
[7] 孙海生, 韩红. AuthorRank 引用认同用于科研人员评价的实证分析[J]. 情报杂志,2011, 30(7):30 -33.
[8] 邓启平,王小梅. 利用 LeaderRank 识别有影响力的作者[J]. 现代图书情报技术. 2015,31(9):60 -67.
[9] 寇晶晶. 中国学者影响力评价进展研究[J]. 情报探索, 2017, 1(11):129 -134.
[10] H - index and variants[EB/OL]. [2017 -08 -22]. http://sci2s. ugr. es/hindex.
[11] 马瑞敏, 韩小林. 基于特征因子算法改进的作者影响力评价研究[J]. 重庆大学学报(社会科学版), 2015, 21(2):106 -109.
[12] MCCORMACK N. Peer review and legal publishing: what law librarians need to know about open, single-blind, and double blind reviewing[J]. Law library journal,2009,101(1): 59 -70.
[13] MULLIGAN A. Quality, certification and peer review[J]. Information Services & Use,2008,28(3 -4):197 -214.
[14] SHOTTON D. The five stars of online journal articles-a framework for article evaluation [EB/OL]. [2017 -09 -13]. http://www. dlib. org/dlib/january12/shotton/01shotton. html.
[15] PERAKAKIS P, TAYLOR M, MAZZA M, et al. Understanding the role of open peer review and dynamic academic articles[J]. Scientometrics, 2011, 88(2):669 -673.
[16] 刘春丽. 基于软同行评议的科学论文影响力评价方法——F1000 因子[J]. 中国科技期刊研究, 2012,23(3):383 -386.
[17] 姜春林,张立伟,刘学. 中外同行评议研究现状及问题探讨[J]. 中国科技期刊研究,2015(3):163 -166.
[18] FORD E. Defining and characterizing open peer review: a review of the literature[J]. Journal of scholarly publishing,2013,44(4) : 311 -326.
[19] 宋丽萍,王建芳. 基于 F1000 与 WoS 的同行评议与文献计量相关性研究[J]. 中国图书馆学报,2012 (2): 62 -69.
[20] 沈小玲,徐勇,严卫中. 基于网络文献计量的科技论文学术影响力综合评价研究[J]. 图书情报工作,2013, 57(21): 95 -103.
[21] F1000 Prime-faqs [EB/OL]. [2017 -08 -22]. https://f1000. com/prime/about/faqs.
[22] F1000 Prime -classifications [EB/OL]. [2017 - 08 - 22]. https://f1000. com/prime/about/whatis/how.
[23] LOLLE S J, VICTOR J L, YOUNG J M, et al. Genome-wide non-mendelian inheritance of extra-genomic informationin arabidopsis [J]. Nature, 2005, 434(7032): 505 -509.

作者贡献说明:
袁国华:设计并实现技术方案、数据分析,撰写和修改论文;
寇晶晶:提出研究思路,设计研究方案,撰写和修改论文;
张建勇:进行论文写作指导;
韩正琪:负责数据清理和论文修改。

The Research of Scholar Influence Evaluation Based on Open Peer Review: Take the F1000 as an Example

Yuan Guohua[1,2] Kou Jingjing[1,2] Zhang Jianyong[1] Han Zhengqi[1,2]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] University of Chinese Academy of Sciences, Beijing 100049

Abstract: [**Purpose/significance**] This paper designed and built an evaluation method of scholar influence based on open peer review. [**Method/process**] First the paper investigated the evaluation methods of scholar influence home and abroad. Then an evaluation method of scholar influence based on open peer review was designed and built. And Empirical research and comparative analysis were conducted. [**Result/conclusion**] The method of scholar evaluation based on open peer review is a new angle of evaluation, which achieves a multi-angle evaluation of scholars and increases the fullness and discernment of scholar influence. The analysis shows that there is no significant correlation between this method and the literature evaluation method, which can be used as a supplement to traditional evaluation methods. The combination of traditional quantitative analysis methods and open peer review methods will be the development trend of future academic evaluation.

Keywords: scholar influence evaluation open peer review F1000

chinaXiv:202308.00646v1